

# Approximating 1-dimensional TSP requires $\Omega(n \log n)$ comparisons

March 28, 2013

## Abstract

We give a short proof that any comparison-based  $(n^{1-\epsilon})$ -approximation algorithm for the 1-dimensional Traveling Salesman Problem (TSP) requires  $\Omega(n \log n)$  comparisons.

## Introduction

In September of 2012 David Eppstein asked the following question on [cstheory.stackexchange.com](http://cstheory.stackexchange.com):

“The one-dimensional traveling salesperson path problem is, obviously, the same thing as sorting, and so can be solved exactly by comparisons in  $O(n \log n)$  time, but it is formulated in such a way that approximation as well as exact solution makes sense. In a model of computation in which the inputs are real numbers, and rounding to integers is possible, it’s easy to approximate to within a  $1 + O(n^{-c})$  factor, for any constant  $c$ , in time  $O(n)$ : find the min and max, round everything to a number within distance  $(\max - \min)n^{-(c+1)}$  of its original value, and then use radix sort. But models with rounding have problematic complexity theory and this led me to wonder, what about weaker models of computation?

So, how accurately can the one-dimensional TSP be approximated, in a linear comparison tree model of computation (each comparison node tests the sign of a linear function of the input values), by an algorithm whose time complexity is  $o(n \log n)$ ? The same rounding method allows any approximation ratio of the form  $n^{1-o(1)}$  to be achieved (by using binary searches to do the rounding, and rounding much more coarsely to make it fast enough). But is it possible to achieve even an approximation ratio like  $O(n^{1-\epsilon})$  for some  $\epsilon > 0$ ?” [2]

The purpose of this short note is to document, in a citable form, a negative answer.

**Theorem 1** *For any constant  $\epsilon > 0$ , every comparison-based  $n^{1-\epsilon}$ -approximation algorithm for 1-dimensional TSP requires  $\Omega(\epsilon n \log n)$  comparisons in the worst case.*

A “comparison-based” algorithm is one that queries the input only with binary (True/False) queries.

Note: after the initial draft of this note, it was pointed out to me that Das, Kapoor and Smid proved the following result in 1997:

**Theorem 2 ([1], Thm. 1)** *Let  $d \geq 1$  be an integer constant. In the algebraic computation-tree model, any algorithm that, given a set  $S$  of  $n$  points in  $\mathbb{R}^d$  and a sufficiently large real number  $r < n$ , computes an  $r$ -approximate TSP-tour for  $S$ , takes  $\Omega(n \log(n/r))$  time in the worst case.*

They also give related results for minimum-spanning trees.

Note that, by taking  $r = n^{1-\epsilon}$ , their result essentially implies Thm. 1. The proof in [1] uses a Ben-Or-style proof, different from the direct proof here.

## Proof of Theorem 1

Fix any constant  $\epsilon > 0$  and an arbitrarily large positive integer  $n$ .

Consider just the  $n!$  "permutation" input instances  $(x_1, x_2, \dots, x_n)$  that are permutations of  $[n] = \{1, 2, \dots, n\}$ . The optimum solution for any such instance has cost  $n - 1$ .

Define the *cost* of a permutation  $\pi$  to be  $c(\pi) = \sum_i |\pi(i+1) - \pi(i)|$ . Model the algorithm as taking as input a permutation  $\pi$ , returning some permutation  $\pi'$ , and paying cost  $d(\pi, \pi') = c(\pi' \circ \pi)$ .

Define  $C = C(n, \epsilon)$  to be the minimum number of comparisons for any comparison-based algorithm to achieve competitive ratio  $n^{1-\epsilon}$  on these instances. Since  $\text{opt}$  is  $n - 1$ , the algorithm must guarantee cost at most  $n^{2-\epsilon}$ . We will show  $C \geq \Omega(\epsilon n \log n) - O(n)$ .

Define  $P = P(n, \epsilon)$  to be, for any possible output  $\pi'$ , the fraction of possible inputs for which output  $\pi'$  would achieve cost at most  $n^{2-\epsilon}$ . This fraction is independent of  $\pi'$ .

$P$  also equals the probability that, for a random permutation  $\pi$ , its cost  $c(\pi)$  is at most  $n^{2-\epsilon}$ . (To see why, take  $\pi'$  to be the identity permutation  $I$ . Then  $P$  is the fraction of inputs for which  $d(\pi, I)$  at most  $n^{2-\epsilon}$ , but  $d(\pi, I) = c(\pi)$ .)

**Lemma 1**  $C(n, \epsilon) \geq \log_2 1/P(n, \epsilon)$ .

*Proof.* Fix any algorithm that uses less than  $\log_2 1/P$  comparisons on each of the inputs. The decision tree for the algorithm has depth less than  $\log_2 1/P$ , so there are less than  $1/P$  leaves, and, for some output permutation  $\pi'$ , the algorithm gives  $\pi'$  as output for more than a  $P$  fraction of the inputs. By definition of  $P$ , for at least one such input, the output  $\pi'$  gives cost more than  $n^{2-\epsilon}$ , so the algorithm is not a  $n^{1-\epsilon}$ -approximation algorithm.  $\square$

**Lemma 2**  $P(n, \epsilon) \leq \exp(-\Omega(\epsilon n \log n))$ .

Before we prove Lemma 2, note that the two lemmas together prove the theorem:

$$C \geq \log_2 \frac{1}{P} = \log_2 \exp(\Omega(\epsilon n \log n)) = \Omega(\epsilon n \log n).$$

To finish proving the theorem, we prove Lemma 2.

Let  $\pi$  be a random permutation. Recall that  $P$  equals the probability that its cost  $c(\pi)$  is at most  $n^{2-\epsilon}$ . Say that any pair  $(i, i+1)$  is an *edge* with cost  $|\pi(i+1) - \pi(i)|$ , so  $c(\pi)$  is the sum of the edge costs.

Suppose  $c(\pi) \leq n^{2-\epsilon}$ .

Then, for any  $q > 0$ , at most  $n^{2-\epsilon}/q$  of the edges have cost  $q$  or more. Say that edges of cost less than  $q$  are *cheap*.

Fix  $q = n^{1-\epsilon/2}$ . Substituting and simplifying, at most  $n^{1-\epsilon/2}$  of the edges are not cheap.

Thus, at least  $n - n^{1-\epsilon/2} \geq n/2$  of the edges are cheap. Thus, there is a set  $S$  containing  $n/2$  cheap edges.

**Claim 1** *For any given set  $S$  of  $n/2$  edges, the probability that all edges in  $S$  are cheap is at most  $\exp(-\Omega(\epsilon n \log n))$ .*

Before we prove the claim, note that it implies Lemma 2 as follows. By the claim, and the naive union bound, the probability that any there *exists* such a set  $S$  is at most

$$\begin{aligned} \binom{n}{n/2} \exp(-\Omega(\epsilon n \log n)) &\leq 2^n \exp(-\Omega(\epsilon n \log n)) \\ &\leq \exp(O(n) - \Omega(\epsilon n \log n)) \leq \exp(-\Omega(\epsilon n \log n)). \end{aligned}$$

To complete the proof of the lemma and the theorem, we prove the claim.

Choose  $\pi$  by the following process. Choose  $\pi(1)$  uniformly from  $[n]$ , then choose  $\pi(2)$  uniformly from  $[n] - \{\pi(1)\}$ , then choose  $\pi(3)$  uniformly from  $[n] - \{\pi(1), \pi(2)\}$ , etc.

Consider any edge  $(i, i+1)$  in  $S$ . Consider the time just after  $\pi(i)$  has been chosen, when  $\pi(i+1)$  is about to be chosen. Regardless of the first  $i$  choices (for  $\pi(j)$  for  $j \leq i$ ), there are at least  $n-i$  choices for  $\pi(i+1)$ , and at most  $2n^{1-\epsilon/2}$  of those choices will give the edge  $(i, i+1)$  cost less than  $n^{1-\epsilon/2}$  (making it cheap).

Thus, conditioned on the first  $i$  choices, the probability that the edge is cheap is at most  $\min(1, \frac{2n^{1-\epsilon/2}}{n-i})$ . Thus, the probability that all  $n/2$  edges in  $S$  are cheap is at most

$$\prod_{(i,i+1) \in S} \min\left(1, \frac{2n^{1-\epsilon/2}}{n-i}\right).$$

Since  $|S| \geq n/2$ , there are at least  $n/4$  edges in  $S$  with  $n-i \geq n/4$ . Thus, this product is at most

$$\left(\frac{2n^{1-\epsilon/2}}{n/4}\right)^{n/4} \leq (8n^{-\epsilon/2})^{n/4} = \exp(O(n) - \Omega(\epsilon n \log n)) = \exp(-\Omega(\epsilon n \log n)).$$

This proves the theorem. □

## References

- [1] Gautam Das, Sanjiv Kapoor, and Michiel Smid. On the complexity of approximating euclidean traveling salesman tours and minimum spanning trees. *Algorithmica*, 19(4):447–462, 1997.
- [2] David Eppstein. Approximate 1d tsp with linear comparisons? <http://cstheory.stackexchange.com/questions/12593/approximate-1d-tsp-with-linear-comparison> September 2012.